optimizing a protein could involve selecting a variant with lower functionality than the parental protein if that is desired.

The terms "aptamer" and "nucleic acid antibody" are used herein to refer to a single- or double-stranded polynucleotide that recognizes and binds to a desired target molecule by virtue of its shape. See, e.g., PCT Publication Nos. WO 92/14843, WO 91/19813, and WO 92/05285.

"Conservative residues" are those amino acid residues that have a similar property, such as similar chemistry. Conservative changes can be based, for example, on similar hydrophobicity, similar hydrophilicity, similar charge, similar propensity for adopting a particular secondary structure, similar shape, etc. Conservative substitution tables providing functionally similar amino acids are known in the art. In one scheme, the following six groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Serine (S), Threonine (T);
2) Aspartic acid (D), Glutamic acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

(*see, e.g.*, Creighton, *Proteins* (1984)).

"Amino acid mutations" are substitutions, deletions or insertions in amino acid sequences. For example, if an alanine occurs in an amino acid sequence, the alanine could be substituted to a serine, it could be deleted or another amino acid residue could be inserted on the amino or carboxy side of the residue. Because alanine and serine are members of the same conserved family of amino acids in the scheme described above, such a substitution can be termed a "conservative substitution." Other schemes can be used.

The term "antibody" as used herein includes antibodies obtained from both polyclonal and monoclonal preparations, as well as: hybrid (chimeric) antibody molecules (see, for example, Winter et al. (1991) *Nature* 349:293-299; and U.S. Patent No. 4,816,567); F(ab')2 and F(ab) fragments; Fv molecules (noncovalent heterodimers, see, for example, Inbar et al. (1972) *Proc Natl Acad Sci USA* 69:2659-2662; and Ehrlich et al. (1980) *Biochem* 19:4091-4096);

single-chain Fv molecules (sFv) (see, for example, Huston et al. (1988) *Proc Natl Acad Sci USA* 85:5879-5883); dimeric and trimeric antibody fragment constructs; minibodies (see, e.g., Pack et al. (1992) *Biochem* 31:1579-1584; Cumber et al. (1992) *J Immunology* 149B:120-126); humanized antibody molecules (see, for example, Riechmann et al. (1988) *Nature* 332:323-327; Verhoeyan et al. (1988) *Science* 239:1534-1536; and U.K. Patent Publication No. GB 2,276,169, published 21 September 1994); and, any functional fragments obtained from such molecules, wherein such fragments retain specific-binding properties of the parent antibody molecule.

As used herein, the term "monoclonal antibody" refers to an antibody composition having a homogeneous antibody population. The term is not limited regarding the species or source of the antibody, nor is it intended to be limited by the manner in which it is made. Thus, the term encompasses antibodies obtained from murine hybridomas, as well as human monoclonal antibodies obtained using human hybridomas or from murine hybridomas made from mice expression human immunoglobulin chain genes or portions thereof. See, e.g., Cote, et al. *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, 1985, p. 77.

The term "sequence alignment" refers to the result when at least two amino acid sequences are compared for maximum correspondence, as measured using one of the following "sequence comparison algorithms." Optimal alignment of sequences for comparison can be conducted by any technique known or developed in the art, and the invention is not intended to be limited in the alignment technique used. Exemplary alignment methods include the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* **2**:482 (1981), the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* **48**:443 (1970), the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* **85**:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), and by inspection.

The "three dimensional structure" of a protein is also termed the "tertiary structure" or the structure of the protein in three dimensional space. Typically the three dimensional structure of a protein is determined through X-ray crystallography and the coordinates of the atoms of the amino acids determined. The coordinates are then converted through an algorithm into a visual representation of the protein in three dimensional space. From this model, the local "environment" of each residue can be determined and the "solvent

accessibility" or exposure of a residue to the extraprotein space can be determined. In addition, the "proximity of a residue to a site of functionality" or active site and more specifically, the "distance of the α or β carbons of the residue to the site of functionality" can be determined. (For glycine residues, which lack a β carbon, the α carbon can be substituted.) Also from the three dimensional structure of a protein, the residues that "contact with residues of interest" can be determined. These would be residues that are close in three dimensional space and would be expected to form bonds or interactions with the residues of interest. And because of the electron interactions across bonds, residues that contact residues in contact with residues of interest can be investigated for possible mutability. Additionally, molecular modeling can be used to determine the structure, and can be based on a homologous structure or *ab initio*. Energy minimization techniques can also be employed.

Although not dependent on three dimensional space, the "residue chemistry" of each amino acid is influenced by its position in a protein. "Residue chemistry" refers to characteristics that a residue possesses in the context of a protein or by itself. These characteristics include, but are not limited to, polarity, hydrophobicity, net charge, molecular weight, propensity to form a particular secondary structure, and space filling size.

The phrase "probability matrix" refers to a matrix for determining the probability that an amino acid can be substituted with another amino acid. Typically this matrix is in the form of an algorithm that determines the probability of substitution from the amino acid and its position. The individual entries in the matrix give a probability for placing a given amino acid in the preselected reference sequence at that position. The algorithm can be based on maintenance of structure, evolutionary diversity amongst a family of proteins and/or other factors described herein, as well as combinations thereof. The phrase "generating a probability matrix" refers to the process of determining the variable upon which the probability matrix will be based and, if needed, developing the algorithm to determine the substitutions in the matrix. The probability matrix can be "normalized" by setting the probability of a particular substitution in the matrix to "1" and correspondingly adjusting the relative probabilities of the other amino acids. The matrix can be normalized to the substitution most favored at that position by the algorithm, or to the value in the matrix for the wild type residue in the reference sequence at that